



ANALYSE DE DONNÉES

Rapport de projet

Plan

- 1. INTRODUCTION ET CONTEXTE**
- 2. STATISTIQUES DESCRIPTIVES**
- 3. ANALYSE EN COMPOSANTES INDÉPENDANTES (ACI)**
- 4. ANALYSE EN COMPOSANTES PRINCIPALES (ACP)**
- 5. CONCLUSION**

ENSEIGNANT : BRUNO DARDAILLON

MASTER 1 DOP
2021-2022

AUTEURS :

THIOUNE BASSIROU 22008769
BARRY MARIAMA LAMARANA 21910623
NDJIM MOUHAMADOU MANSOUR 21613317
YASSINE EL HARRAB 21909693

1. INTRODUCTION ET CONTEXTE :

Dans le cadre d'un projet visant l'étude des cours d'eau et plus particulièrement la Touques (un fleuve de basse Normandie), il a été relevé et stocké dans une base de données les températures prélevées par 4 sondes posées d'amont en aval le long de la Touques.

Ces sondes sont en général attachées aux racines des arbres de la sylvestre et immergées dans l'eau à une profondeur leur permettant d'enregistrer la température quelle que soit la saison.

Les températures de ces eaux varient naturellement avec les saisons, tandis que l'eau est généralement plus froide au fond (en raison de l'apport d'eau souterraine) et plus chaude à la surface (en raison de la température de l'air plus élevée). Les évolutions de température d'un cours d'eau peuvent être causées par plusieurs activités :

- le rejet d'eau de refroidissement plus chaude des centrales thermiques,
- le rejet d'eau des barrages,
- l'élimination des plantations riveraines qui ombragent et maintiennent les températures dans les cours d'eau,
- une réduction ou une augmentation des niveaux d'eau en raison au captage ou au détournement de l'eau utilisée pour l'irrigation,
- l'ajout d'eau géothermique plus chaude etc.

Ainsi la présente étude a pour objectif d'apporter une analyse sur la variabilité spatiale des températures et examiner les effets des facteurs environnementaux sur leur hétérogénéité.

Au préambule, nous procéderons à la structuration et à la présentation de l'information contenu dans notre base de relevés, forte de 93 832 observations afin d'en tirer des renseignements.

Cette base de données est composée de 6 variables qui sont :

t	le temps, toutes les deux heures
date	la date, en jour, mois et année
Teau	la température de l'eau
id_sonde	la sonde qui a relevé la température. La sonde peut prendre 4 valeurs, 825, 827, 828 et 830
Rainf.EOBS	le volume de pluie issu du site ECAD ¹
Tair.EOBS	la température de l'air issu du site ECAD

Dans la suite, on cherchera principalement à traiter une telle base de relevés de températures pour découvrir, comprendre ou valider factuellement des phénomènes difficilement observables sans traitement de données.

2. STATISTIQUES DESCRIPTIVES :

Commençons dès à présent par une analyse globale de notre base de données et partons du fait qu'elle nous apportera assurément de précieuses informations.

On étudie une base de température relevée entre 2013 et 2018 dans un fleuve de basse Normandie appelée la Touques. Quatre sondes posées d'amont en aval le long de ce fleuve ont été chargés de relevés ces températures. Il s'agit des sondes **825, 827, 828, 830**.

Respectivement de l'ordre selon lequel elles ont été posées, les températures prises par ces sondes augmentent d'amont en aval et les statistiques descriptives pourront nous permettre de justifier ce constat.

La **Figure 1** met en évidence les statistiques élémentaires telles la **moyenne**, la **médiane**, les **minimums** et **maximums**, les **premiers** et **troisièmes** quartiles.

Il est à souligné que ces mesures sont purement informatives. Nous remarquons que :

- La température moyenne récupérée par la sonde **825** est **2,14 %** moins élevée que celle prise par la sonde **827**.
- La température moyenne récupérée par la sonde **827** est **1,61%** moins élevée que celle prise par la sonde **828**.
- La température moyenne récupérée par la sonde **828** est **4,05%** moins élevée que celle prise par la sonde **830**.

On en conclue que plus la sonde est posée en amont et plus la température est fraîche.

	Min	1st Qu	Median	Mean	3rd Qu	Max
Sonde 825	3.799	8.658	11.428	11.262	13.832	17.809
Sonde 827	2.115	8.699	11.673	11.419	14.138	8.545
Sonde 828	1.207	8.606	11.861	11.702	4.847	20.208
Sonde 830	1.897	9.247	12.342	12.302	15.461	21.135

Figure 1

3. ACI : Analyse en Composantes Indépendantes

Une première ACI avec fastICA a permis d'extraire 4 composantes dont une seule ne variait presque pas lors de la décomposition de son signal, elle a été donc mise de côté pour en extraire finalement que les 3 composantes beaucoup plus pertinentes.

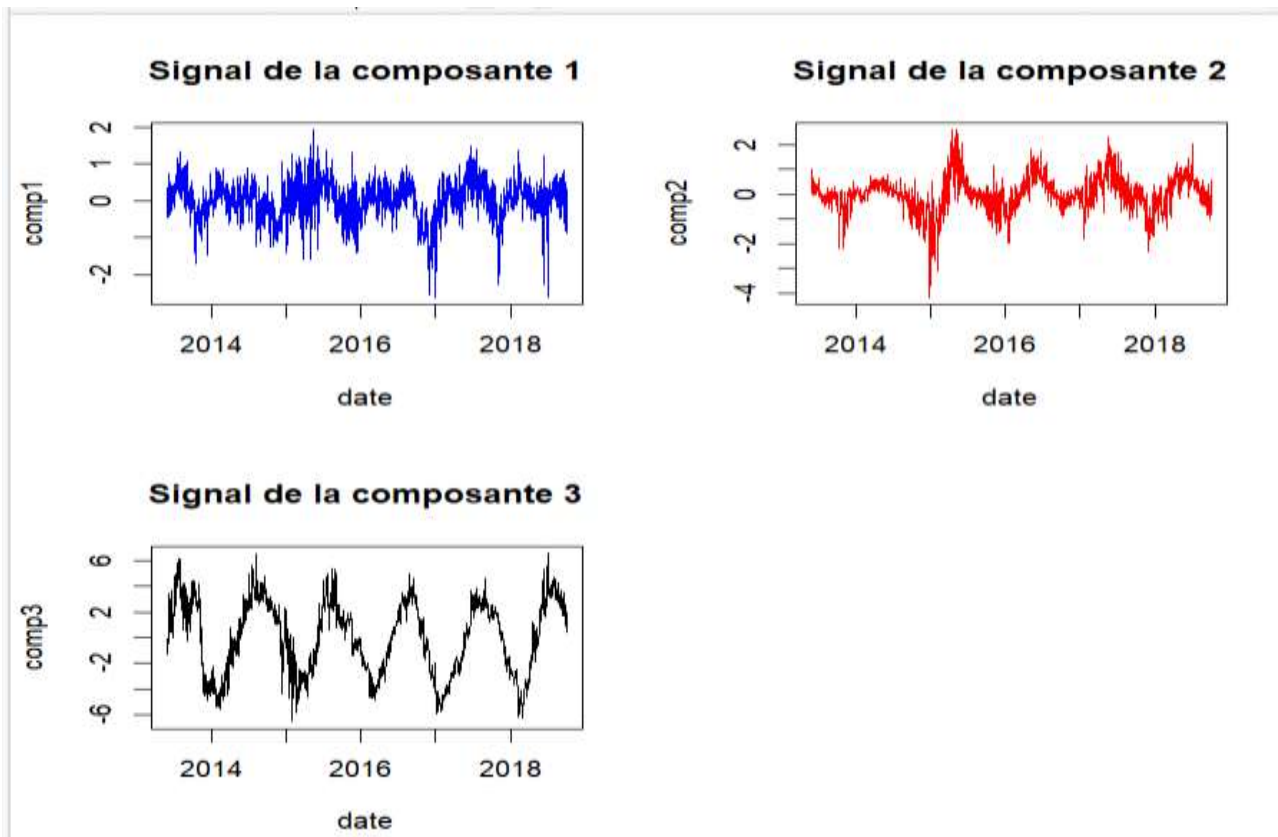


Figure 2

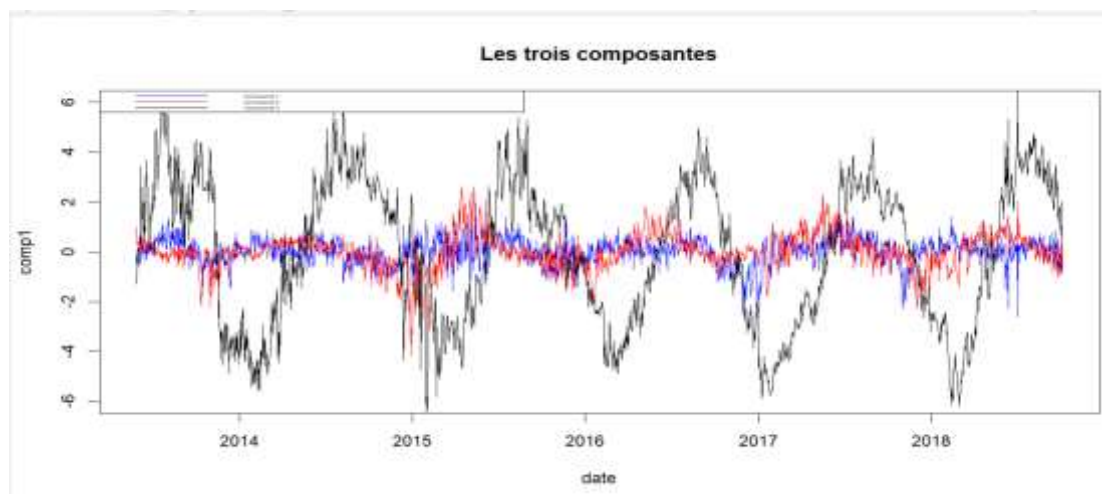


Figure 3

La sortie obtenue à la [Figure 2](#) décrit une saisonnalité de nos composantes. C2 semble décliner le plus par rapport à C1, elle prend pour l'année 2015 des valeurs extrêmes. Ce phénomène est dû peut-être à une cascade d'événements lié à des variations météorologiques brutales. Le déphasage des courbes ([Figure 3](#)) laisse ainsi présager que le signal d'une composante variant en premier serait celui de la température de l'air (courbe bleu) et celui qui le suit (courbe rouge) serait celle de l'eau. Cela semble amissable car l'air se réchauffe relativement vite comparé à l'eau.

4. **ACP** : Analyse en Composantes Principales

La mise en place de l'ACP nous permet de vérifier la corrélation existante entre les variables et de découvrir puis d'expliquer assez nettement nos composantes extraites de l'ACI.

Ainsi, une deuxième ACI avec fastICA nous a permis de décomposer en deux composantes le signal **D** (Tw-Ta) qui signifie la différence entre la température de l'eau et celle de l'air. Ces deux composantes que nous avons nommées **C1** et **C2** vont intervenir en tant que variables actives parmi d'autres variables, c'est-à-dire participeront à la fixation des axes de notre ACP.

Le diagramme des éboullis nous a permis de savoir qu'il serait beaucoup plus intéressant de travailler sur les deux premiers axes (Dim 1 et Dim 2) qui représentent à eux deux 92,6% de l'inertie totale. Ce qui est largement suffisant pour confirmer nos hypothèses précédemment établies.

L'ACP nous permet d'exploiter jusqu'à 5 axes d'interprétation. Mais nous allons nous concentrer que sur deux axes parmi les cinq dont les valeurs absolues sont supérieures à 1. On a donc appliqué le critère de **Kaiser**.

Ainsi, l'ACP proposée (**Figure 4**) résume **92 %** de l'information. Celle-ci est portée par l'axe 1 à **71 %** et par l'axe 2 à **21%**. Il y a 7 variables qui sont utilisées : **5 actives** (C1/C2/Tw/Ta/D) et **2 supplémentaires** (PE/P). Les variables actives participent à la fixation des axes et celles supplémentaires ne contribuent pas à la fixation des axes mais permettent de mieux interpréter notre graphique.

Deux tendances semblent se dégager : la dimension une (**Dim1**) est portée par **Tw, Ta et D(Tw-Ta)** à **74%** tandis que la dimension deux (**Dim2**) est fixée par **C1 et C2** à **95 %**. Ces valeurs nous proviennent du **tableau des contributions** issues de l'ACP.

Certaines variables apparaissent très liées comme Ta et Tw désignant respectivement la température de l'air et de l'eau, la **matrice des corrélations** vient confirmer ce rapprochement car les deux variables ont une corrélation de **0,91**.

Il apparaît bien une corrélation nulle entre les deux composantes C1 et C2 provenant de l'ACI donc indépendantes et dont les faisceaux sont parfaitement orthogonaux dès lors que les deux variables sont bien représentées dans le plan (1,2) à **0,90** et **0,99** respectivement pour C1 et C2 d'après le **tableau des cosinus carrés** désignant les qualités de représentation.

Les Ta et la composante C1 sont opposées. Ces deux variables sont fortement corrélées mais négativement (**-0,78** d'après la matrice des corrélations). Cette négativité traduit le fait que lorsque l'une augmente, l'autre diminue.

Les variables PE et P pourraient ne pas être assez bien représentées car très proches du centre du cercle.

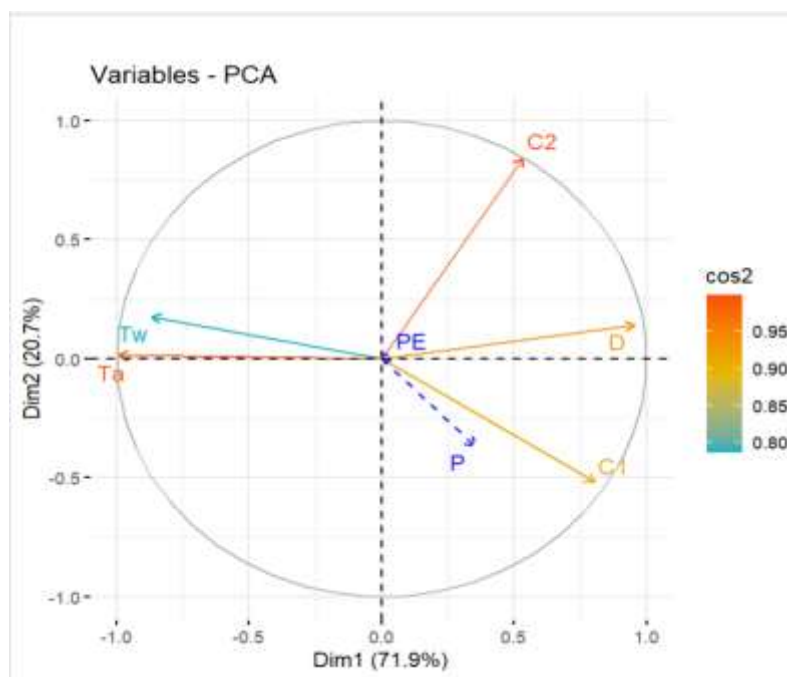


Figure 4 :

5. Conclusion :

En somme, après ce travail de modélisation et d'extraction de connaissances, nous pouvons conclure que les phénomènes météorologiques influent fortement sur la température de l'eau.

Au vu des résultats, il est clair que la température de l'air peut impacter sur celle des cours d'eau de la Touques. Cependant d'autres évènements entrent en jeu dans ce phénomène observé. Ces événements sont : la pluviométrie, la fonte de neige, ruissèlement sur sol, ensoleillement. Il est à noter également que les activités humaines sur l'environnement (rejet d'eaux usées, émission de polluants) ont des répercussions directes sur la variation de la température des cours d'eau en effet la perturbation de l'équilibre vitale des êtres vivant sous ces eaux.